

Article

# Images of Roman Imperial Denarii: A Curated Data Set for the Evaluation of Computer Vision Algorithms Applied to Ancient Numismatics, and an Overview of Challenges in the Field

Ognjen Arandjelović \*  and Marios Zachariou

School of Computer Science, University of St. Andrews, St Andrews KY15 5BG, UK;  
marios.zachariou@hotmail.com

\* Correspondence: ognjen.arandjelovic@gmail.com; Tel.: +44-(0)1334-48-26-24

Received: 8 March 2020; Accepted: 12 March 2020;

First Version Published: 16 March 2020 (doi:10.3390/sci2010015)

Second Version Published: 14 June 2020 (doi:10.3390/sci2020047)



**Abstract:** Automatic ancient Roman coin analysis only recently emerged as a topic of computer science research. Nevertheless, owing to its ever-increasing popularity, the field is already reaching a certain degree of maturity, as witnessed by a substantial publication output in the last decade. At the same time, it is becoming evident that research progress is being limited by a somewhat veering direction of effort and the lack of a coherent framework which facilitates the acquisition and dissemination of robust, repeatable, and rigorous evidence. Thus, in the present article, we seek to address several associated challenges. To start with, (i) we provide a first overview and discussion of different challenges in the field, some of which have been scarcely investigated to date, and others which have hitherto been unrecognized and unaddressed. Secondly, (ii) we introduce the first data set, carefully curated and collected for the purpose of facilitating methodological evaluation of algorithms and, specifically, the effects of coin preservation grades on the performance of automatic methods. Indeed, until now, only one published work at all recognized the need for this kind of analysis, which, to any numismatist, would be a trivially obvious fact. We also discuss a wide range of considerations which had to be taken into account in collecting this corpus, explain our decisions, and describe its content in detail. Briefly, the data set comprises 100 different coin issues, all with multiple examples in Fine, Very Fine, and Extremely Fine conditions, giving a total of over 650 different specimens. These correspond to 44 issuing authorities and span the time period of approximately 300 years (from 27 BC until 244 AD). In summary, the present article should be an invaluable resource to researchers in the field, and we encourage the community to adopt the collected corpus, freely available for research purposes, as a standard evaluation benchmark.

**Keywords:** coins; review; problems; data corpus; grade; preservation; condition

## 1. Introduction

It is no longer an exaggeration to say that computer vision is pervasive in everyday life: Face detection [1,2] has been a standard feature of digital cameras and smartphones for well over a decade, online image depositories are increasingly successful at categorizing images by their semantic content (scene: Beach, city, countryside, etc; objects: Cars, buildings, dogs, churches, statues, etc.) [3,4], automatic diagnosis and prognosis of diseases has even surpassed the performance of human experts in some domains [5–7], etc. This success, coupled with the increasing pervasiveness of powerful computing devices and the dramatic improvement in user-friendliness of technology in general, is having a positive impact on inter-disciplinary research, with a growing interest in the application

of modern computer science in other scientific fields, as well as in the arts and humanities [8–10]. A particularly interesting domain of application concerns ancient numismatics, i.e., the study of ancient currency, which has been attracting an increasing amount of attention from the computer vision community. The focus of the present article is on a number of mainly methodological issues that are important in this increasingly prolific research area, which we argue have received insufficient attention in the published literature to date. To understand our contributions, it is necessary to introduce some basic numismatic terminology, which we do next.

## 2. Computer Vision and Machine Learning Challenges within the Domain of Ancient Numismatics

We begin this section with an explanation of the relevant numismatic terminology necessary for the understanding of the present article and the related literature, then categorize and describe in detail the most important (practically and technically) challenges in the field, and summarize the progress to date in addressing these.

### 2.1. Terminology

The specialist vocabulary of numismatics is extremely rich, and its comprehensive review is beyond the scope of the present article [11]. Herein, we introduce a few basic concepts that are important for the understanding of the present contribution and the related works.

Firstly, when referring to a ‘coin’, the reference is being made to a specific object, a physical artifact. It is important not to confuse it with the concept of a (coin) ‘issue’, which is more abstract in nature [12]. Two coins are of the same issue if the semantic content of their obverses and reverses (heads and tails in modern, colloquial English) is the same. For example, if the obverses show individuals (e.g., emperors), they have to be the same individuals, be shown from the same angle, have identical headwear (none, crown, wreath, etc.), be wearing the same clothing (drapery, cuirass, etc.), and so on. Moreover, any inscriptions, usually running along the coin edge (referred to as the ‘legend’), also have to be identical, though not necessarily be identically arranged spatially letter by letter [13]. Online Coins of the Roman Empire (OCRE; see <http://numismatics.org/ocre/>), a joint project of the American Numismatic Society and the Institute for the Study of the Ancient World at New York University, lists 43,000 published issues. The true count is likely to be even greater.

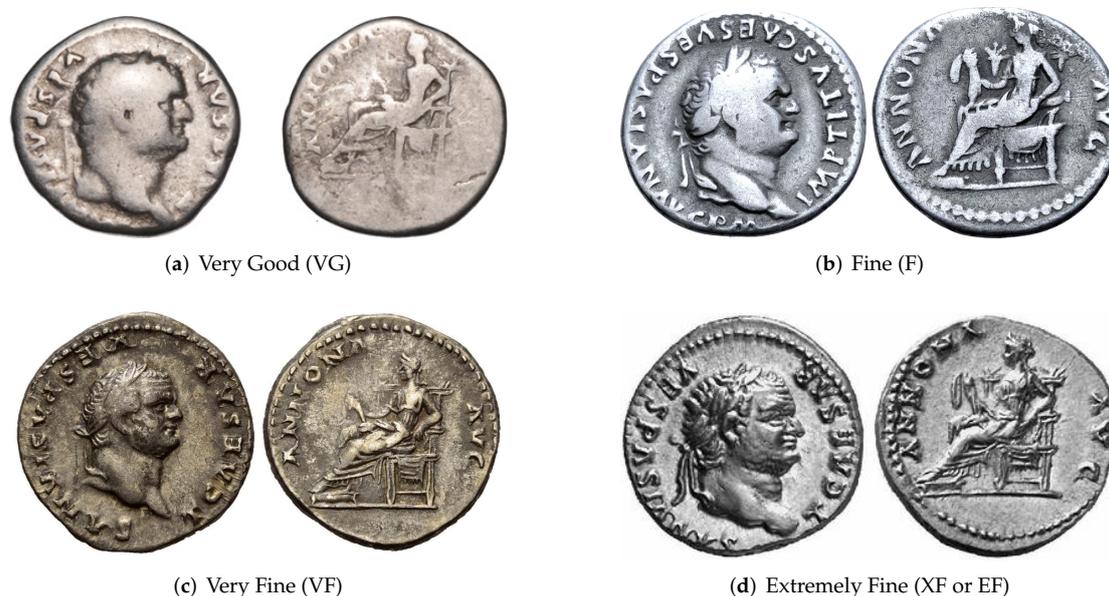
### 2.2. Grading

An important consideration in numismatics regards the condition of a particular coin. As objects that are a millennium and a half to three millennia old, it is unsurprising that, in virtually all cases, they have suffered damage. This damage was effected by a variety of causes. First and foremost, as most coins were used for day-to-day transactions, damage came through proverbial wear and tear. Damage was also effected by the environment in which coins were stored, hidden, or lost, before being found or excavated—for example, the moisture or acidity of soil can have significant effects. Others were intentionally modified, for example, for use in decorative jewellery.

The amount of damage to a coin is of major significance both to academic and hobby numismatists. To the former, the completeness of available information on rare coins is inherently valuable, but equally, when damaged, the type of damage sustained by a coin can provide contextual information of the sort discussed earlier. For hobby numismatists, the significance of damage is twofold. Firstly, a better-preserved coin is simply more aesthetically pleasing. Secondly, the price of the coin, and thus its affordability as well as its investment potential, are greatly affected: The cost of the same issue can vary by 1–2 orders of magnitude.

To characterize the amount of damage to a coin due to wear and tear, as the most common type of damage, a quasi-objective grading system is widely used. Fair (Fr) condition describes a coin so worn that even the largest major elements are mostly destroyed, making even a broad categorization of the coin difficult. Coins of Very Good (VG) grade have most detail worn nearly smooth around the central areas but still visible on the periphery. Fine (F) condition coins show significant wear with many minor

details worn through, but the major elements are still clear at all of the highest surfaces. Very Fine (VF) coins show wear to minor details, but clear major design elements. Finally, Extremely Fine (XF) coins show only minor wear to the finest details. Examples are shown in Figure 1.



**Figure 1.** Examples of the same coin issue (denarius of emperor Titus; RIC 972 [Vespasian], RSC 17, BMC 319) in different grades of conservation: (a) Very good (VG), (b) fine (F), (c) very fine (VF), and (d) extremely fine (XF or EF). The two lowest grades, namely fair (Fr) and good (G), are not shown due to the lack of interest in specimens damaged so severely.

### 2.3. Practical Applications

One of the features of numismatics which makes it an interesting domain for the application of computer vision and machine learning lies in the number and diversity of specific problems that it presents. Many of these directly correspond to challenges faced by experts or hobby collectors, though some new work introduces innovative challenges which are only possible with the use of technology (we shall elaborate on this shortly). The key problems, few of which can be considered anywhere near solved, include the following:

- Specimen matching,
- Issue matching [14–16],
- Denomination categorization,
- Issuing authority recognition [17],
- Legend readout [13],
- Semantic analysis [18],
- Forgery recognition, and
- Die matching.

As implicitly explained in the previous section, *specimen matching* refers to the problem of determining if the same coin specimen in two images is the same, i.e., if they show the same actual physical artifact. There are several important applications of this task. For example, it can be used to determine the provenance of a specific coin or to track its value across time as it is sold and passed on from one collector onto another. Importantly, specimen matching can also be used to automatically monitor massive volumes of coins sold on non-traditional auction web sites, such as eBay, and to track stolen coins. The key challenges for specimen matching lie in differential appearance effected by

different illumination conditions, camera settings (e.g., aperture, focus, and exposure), clutter, scale, and viewpoint [19].

In contrast to specimen matching, *issue matching* refers to the problem of determining if the coins shown in two images are of the same issue, i.e., if they contain the same semantic content and are of the same denomination (e.g., denarius, anotoninianus, follis, sestertius). This task is the first and probably the most commonly performed one by any numismatist; colloquially put, it answers the question “What is this coin I’ve got?”. In addition to all of the aforementioned challenges outlined in the context of specimen matching, in issue matching, a major challenge of a semantic nature emerges: Recall that issues are identified by the corresponding semantic contents, which can exhibit both stylistic variability (e.g., due to different die engravers), appearance change due to physical damage or chemicals in the environment, or die wear, to name but a few; see Figure 2. Recalling from the previous section that the number of different issues of Roman Imperial coins exceeds 43,000, it is not difficult to see why issue matching is inherently an extremely difficult problem [15]. In addition, such a high number of classes makes it all but practically impossible to obtain an annotated gallery of exemplars of all (or most) issues [18,20].



**Figure 2.** Reverses of two different specimens of the same issue—a silver (AR) denarius of Julia Maesa (RIC 249). Despite them being the same issue, the two specimens exhibit a series of appearance differences. These range from the arrangement of legend letters (e.g., note that the ‘T’ in ‘FECVNDITAS’ is to the left—as seen by a reader—of the goddess depicted on the specimen in (a) and to the right in (b)), the exact pose of the child next to the goddess (Feccunditas) or indeed the goddess herself, the flan shape and the centering of the motif within the flan, the damage and loss of fine detail, and the toning (‘color’ change).

*Denomination categorization* is a classification problem which, as the name suggests, is concerned with the determination of the denomination of a coin. Denarii, antoniniani, sestertii, ases, and dupondii are examples of the most common denominations of the Roman Imperial period before the economic crisis of the third century. Some of these are shown in Figure 3. The knowledge of a coin’s denomination can be useful as a step aiding in issue matching or in its own right for monitoring market trends (types of coins being sold, price changes, etc.).

Most Roman imperial coins feature a portrait (all but universally in profile, and usually facing right). Most often, this is the current emperor, sometimes their predecessor (as commemoration following their death), and also frequently their spouse. The recognition of this individual is one of the first things that a numismatist will do in the process of identifying a coin, i.e., it is a step in the process of issue recognition. Within the scope of computer-vision-based analysis of ancient coins, *issuing authority recognition* started attracting attention following the realization that tackling issue recognition is a far more difficult challenge than anticipated at first. Hence, the attempts to apply generic object recognition algorithms waned in popularity, and instead, the focus shifted towards the use of more

domain-specific knowledge, the recognition of the depicted person being an obvious choice. Thus, the challenge of *legend readout* concerns the recognition of the legend inscription. So far, it has received little attention from the computer vision community [13] despite its utility to numismatists. In large part, this is likely a consequence of the difficulty of the problem: Legends are abounding in fine detail and are prone to damage, with letters easily confused with one another, or indeed a damaged letter with a legend break.



**Figure 3.** Coins of different denominations of the same emperor, Domitian. Shown are, in order of value at the time of their use, examples of an (a) as, (b) dupondius, (c) sestertius, (d) denarius, and (e) aureus.

The legend on an ancient Roman imperial coin is an interesting semantic element. Some parts of it contain, in essence, the same information as the motif they encircle. For example, on the obverse, the legend almost invariably explicitly names the issuing authority shown on the coin—in Figure 4a, it begins with the 'AVRELIVS', which refers to Marcus Aurelius. Thus, this information can be used to aid in the process of issuing authority recognition or, with reference to the reverse, in the interpretation of the corresponding motif. However, the legend also contains some information which is generally not contained elsewhere. For example, the legend often contains the consular year of the issuing authority, such as 'COS III' (third consular year), which allows for the precise dating of the issue and its disambiguation from other issues otherwise identical to it.

We have already discussed issue matching as probably the most important and pervasive problem in automatic ancient coin analysis. A major and indeed fundamental problem with the existing approaches which rely on visual matching of images, as highlighted in Section 2.1, is that the number of classes in this classification problem is enormous, exceeding 43,000. This is not only a technical challenge, but also a practical one: It is virtually impossible to obtain gallery samples of such a high number of issues or indeed anything even close to that number. Yet, this was only recently explicitly recognized in the literature [18]. Thus, recently, an alternative approach was first put forward, as well

as the first promising steps towards its implementation. The idea is very much akin to what a human numismatist does: Interpret and *understand* the semantic content [21] of a coin (hence, *semantic analysis*), and then use this semantic description for matching against textual reference entries [22]. Thus, the visual matching problem is eventually turned into a text-matching one. This work is still in its early stages, but highly promising results have already been reported using a deep-learning-based framework capable of automatically learning salient concepts and the range of their artistic depiction variability [20].



**Figure 4.** Examples of appearance variability in the depiction of the same individual (emperor Marcus Aurelius) on different coin issues (a–d). The challenge of ‘face recognition’ in this context, involving artistic stylization and abstraction, can be seen to eclipse that of conventional face recognition—a problem which has been attracting an enormous amount of research attention for some five decades, and yet still remains unsolved for nearly all practical purposes.

Considering the size of the global ancient coin market, it is hardly surprising that it is an attractive target for fraud. Unlike most other ancient artifacts (e.g., highly ornate pottery, helmets and other armor, swords, etc.), for the most part, ancient coins are medium-value collectables. This makes it cost-ineffective to individually authenticate all but a small number of more expensive specimens. Yet, the high volume of sales makes forgery a lucrative business. Despite this major practical significance, interestingly, the task of automatic *forgery detection* has not been explored in any published work to date. What makes this observation even more surprising is that the problem is technically quite interesting. In particular, the novel challenge lies in the new kind of intra-class variability within the class of forgeries. This variability emerges as a consequence of different methods used to produce fake coins. While a thorough discussion of this topic is beyond the scope of the present article, the simple example in Figure 5 will serve to illustrate the gist of it. Specifically, compare an authentic example of a silver denarius of Clodius Albinus in Figure 5a with the three forgeries in Figure 5b–d. The first of the latter, in Figure 5b, is good in style, and was likely produced from a casting mold, itself made from an authentic specimen (as a ‘negative’ thereof). The lack of authenticity is given away by the casting sprue at 10–11 o’clock looking at the obverse, the relief pattern around the legend (especially on the reverse; it is highly unlike that of struck coins), and the surface of the coin (impressions of small casting bubbles). In contrast, the forgeries in Figure 5b,c are poor in style, mostly likely made from modern molds, and readily recognizable as being produced by casting and not striking. How this wide inter-class variability can be learned is an open question and arguably makes the problem one of novelty detection.

Recall that ancient Roman imperial coins were minted by striking a blank coin placed between hand-carved dies [23]. This is in contrast to casting, which was used briefly during the Republican period, as well as later in the production of medallions, probably due to their much larger size (often in excess of 50 g). Being able to tell if two coin specimens of the same issue were made using the same dies, i.e., *die matching*, is of much interest to research numismatists (and much less so to hobby collectors), because, for example, this allows for the inference of migratory patterns of peoples, trading routes, etc. Die matching can also assist in the fight against high-quality forgeries, some of which were

struck in modern times but using ancient dies or copies thereof [24]. To the best of our knowledge, die matching remains an entirely unexplored challenge in the realm of automated ancient coin analysis.



**Figure 5.** Example of (a) an authentic silver denarius of Clodius Albinus, and (b–d) three examples of forgeries (different issues of silver denarii, all with Clodius Albinus as the issuing authority). Though reasonably convincing at first sight, the forgery in (b) is in fact a cast (as witnessed by the sprue at 10–11 o’clock looking at the obverse, as well as the fine surface features). The forgery in (c) is poor in style and thus utterly unconvincing. The poor style (though less so than the previous example) and inappropriate metal composition, evident from the toning of the specimen, also make the forgery in (d) an unconvincing one.

#### 2.4. Research Effort to Date

As noted in the previous section, most research on the application of computer vision and machine learning in the domain of ancient numismatics focused on the problem of issue recognition, or, more specifically, visual issue matching. Within this body of work, in terms of technical underpinnings, visual matching based on local features (chiefly SIFT [25]) dominates the literature [14,15,26]. Though highly successful in a wide variety of object recognition tasks [27–30], these approaches were quickly found to perform very poorly in the context of the problems of interest herein, showing some success only in highly controlled conditions, i.e., when changes in illumination are small or non-existent, when images are devoid of clutter, and when the coins are canonically oriented. This is highly unrealistic in practice: Assumptions of limited photometric variability do not hold, and the removal of clutter (segmentation) is difficult, as is geometric registration [19]. An illustration of just some of the challenges is shown in Figure 6.

In hindsight, the disappointing performance of local-feature-based methods ought not to be surprising. Firstly, ancient coins do not possess discriminative textural information [31,32]. Textural variability is a confounding factor. Rather, appearance variation emerges from geometry (3D) of coins and, thus, the manner in which light is reflected off them. Thus, in terms of local appearance, most coins look alike—the absence of the use of their geometric relationships is crucial. Driven by this insight, the best-performing local-feature-based method builds compound features in the form of directional histograms centered at automatically detected interest points [14]. Thus, both local and distal appearances are integrated, and the geometric relationship is captured. Nevertheless, though significantly surpassing the performance of the existing method at the time, even this method failed to demonstrate practically useful matching rates.

Driven in part by the lack of success of what may be termed ‘conventional computer vision’ approaches on the one hand and the groundbreaking achievements of deep-learning-based methods on the other, much like other recent cultural-heritage-focused computer science work [33–35], more recent efforts in automatic ancient coin analysis have turned their attention to the use of neural networks. Thus, Schlag and Arandjelović [17] proposed a VGG16 deep-neural-network-based algorithm for issuing authority recognition, and demonstrated outstanding performance on three large corpora of data. Aslan et al. [16] used a pre-trained ImageNet, adapted to the domain using transfer learning, on a small data set of Roman republican coins with lesser success. The deep learning algorithms of Cooper and Arandjelović [18,20] and Anwar et al. [36] both focus on the semantics of motifs depicted on coins, the former on Roman imperial and the latter on Roman republican coins—the problem which we already noted as being extremely promising in terms of practical significance, and most interesting from the technical viewpoint.



**Figure 6.** Most ancient coins are sold by private individuals—the non-professional imaging setup used introduces a series of additional confounds and challenges, such as background clutter, poor illumination, partial occlusion, etc. as illustrated on real-world examples (a–d).

Related work, not falling under the umbrella of computer vision as such nor machine learning, includes the acquisition of 3D scans of ancient coins [37–39]. This body of research is closer in

spirit to efforts on the digitization and visualization of cultural artifacts [40,41], including temporal modeling [42–44] and hyperspectral imaging [45–47].

### 3. Curation: Motivation Thereof and Our Data

A major limitation of the published work in the realm of computer vision and machine-learning-based analysis of coins, or rather, of the evaluation methodology of this body of work, lies in the absence of an understanding of the heterogeneity of data used in it. In Section 2.2, we introduced the common standard for quantifying the degree of wear and tear suffered by a particular specimen (often referred to as the condition of the coin). Even the most inexperienced of ancient numismatists understands that, in general, the condition of a coin greatly affects the scope of analysis that it is useful for. As noted in Section 2.3, in certain instances, the legend is necessary for the precise determination of a coin's issue; yet, the legend, containing fine detail and some of the more elevated detail surfaces, often gets damaged significantly.

The aforementioned issues are likely to be even more significant when automatic computer-based analysis is used. Despite that, this problem was not recognized until the work of Fare and Arandjelović [48], who were the first to bring it to attention. In part, this is likely a consequence of the difficulty of obtaining a curated data set; hence, our present effort and contribution.

The earliest work generally used coins of a very high grade (in about Extremely Fine condition, and, notably, a small number of issues) [15,26,49]. Not only does this limit the scope of insights which the corresponding experiments provide, coins such as these are of the least interest in the context of automated analysis. Firstly, these coins are rare and comprise a very small proportion of coins handled by most numismatists. There is little gain in automating their processing. Secondly, exactly because coins in such a high state of preservation are rare, they are usually rather expensive and are sold by specialist dealers, and therefore normally accompanied with detailed information already (often including their provenance, previous owners and sales, etc.). Coins like these are normally not accidentally stumbled upon. The first work that included a more representative sample of real-world data is that of Arandjelović [14], who also used a much larger corpus (circa 3000 specimens). At the same time, because the corpus was not labeled according to the condition, it is difficult to gain much insight into the behavior of the proposed algorithm (or indeed any algorithm evaluated on the same data) and to seek an understanding of how well it performs as a function of a query (or gallery) coin's grade. The work which followed [17,18,20] also used larger and more diverse corpora, but again without any condition-based stratification.

#### 3.1. Our Corpus of Roman Imperial Denarii

Considering the concerns and limitations that we identified and discussed related to the data sets used in the existing published literature, we carefully considered a series of issues in collecting and curating the data set introduced herein. We discuss each of these next, and conclude with a summary description of our corpus.

##### 3.1.1. Why Denarii?

As summarized in Section 2.3, there were a range of different denominations used in Imperial Rome during its existence (i.e., from 27 BC, when Augustus became the first emperor, until the fall of Rome in 476 AD). Recalling our aim of collecting a data corpus curated by the condition of coins it contains, there are several reasons for why we decided to focus on denarii in particular.

Firstly, the denominations, such as the dupondii, ases, and sestertii, featured rather large and heavy coins. Sestertii, for example, usually weigh between 25 and 28 g, and have a diameter of 32–34 mm. Being of lower value, these coins were also extensively used. For these reasons, they normally suffered significant damage, which makes it difficult to source a sufficient number of examples in different states of preservation. Moreover, the aforementioned denominations were all made of more reactive materials (copper alloys) and, as a result, experience color variability due to

reactions with environmental agents (moisture, acids, etc.), thus possibly introducing undesirable confounding factors to the data. Lastly, the use of these denominations declined over time, limiting the range of motifs and styles to a specific period of the Empire, and thus potentially creating a bias in the data.

On the other hand, aureii are extremely rare and difficult to find in the required range of grades—being extremely valuable (25 denarii, or about five weeks' salary of a Roman soldier), they were not used in general circulation and are usually very well preserved. More obscure denominations, such as semises and quadrantes, are also extremely rare, were used only during a short period, and exhibit the same issues related to the alloys they were made of as dupondii, ases, and sestertii, discussed earlier. Lastly, although common, antoniniani were issued over a period of only six decades, limiting the range of issuing authorities which could be covered, as well as reverse motifs and artistic styles.

Denarii, on the other hand, do not exhibit any of the aforementioned limitations. They were made of comparatively non-reactive silver (confusingly, abbreviated to AR in numismatics, and not Ag as in chemistry) and thus experience little to no discoloration (so-called toning changes are perceptually equivalent to a simple darkening of the surface), were used extensively from the beginning of the empire until the economic crisis of the third century (i.e., for about 300 years), exhibit a wide variability in motifs depicted on them, and were common enough that a diverse set of grades is not overly difficult to find.

### 3.2. Curation

In Section 2.2, we summarized the common standard for grading ancient coins. It is important to note that the descriptions of different grades leave room for some subjectivity—hence, our use of the term 'quasi-objective'. In the context of the present work, the practical significance of this observation lies in the potential difficulty of ensuring accuracy. To ensure that this goal is met, it is imperative that the grading is performed by an expert and that a sufficient number of graders are used so as to average out any potential biases. Thus, all of the data in our corpus are obtained from reputable dealers and auction houses, as are the accompanying labels. An attractive feature of this approach also lies in the self-regulating control of bias—any systematic bias (e.g., overestimation of the state of preservation) would end up being self-defeating, as it would reduce the incentive to purchase the bulk of the coins which are not of the top grade.

As regards the choice of grades sought for inclusion, we focused our attention on three: Fine, Very Fine, and Extremely Fine. This choice was a simple one and was motivated firstly by the observation that coins in a condition worse than Fine are seldom of interest to either scholars or hobby collectors; the exceptions are invariably extremely rare issues. The 'collectable' range, which includes coins in Fine condition or better, is widely relevant to numismatists on the one hand and exhibits major appearance variation (even in the absence of other confounding factors), thus making it of interest to computer scientists.

Indeed, a major difficulty in collecting the present data set was discovered to emerge precisely in the breadth of conditions sought—denarii, which were expensive enough to be sold by reputable auction houses in Fine condition, are usually far too rare in Extremely Fine condition, whereas those which are readily found in Extremely Fine condition were seldom seen in Fine condition due to their low cost. Nevertheless, domain expertise helped us to direct our search, and we were successful in collecting 100 different issues, each with multiple specimens in all three grades: Fine, Very Fine, and Extremely Fine.

### 3.3. Data Set Description

Having discussed the reasons underlying our curation criteria and the choices ultimately made, what is left is to describe the data set we collected. We also note that this data set is available freely for research purposes upon request from the corresponding author.

Our corpus of Roman imperial denarii comprises 100 different coin issues, each with multiple examples of specimens in Fine, Very Fine, and Extremely Fine conditions. Each issue usually contains two examples (different specimens) for each condition, and some contain more, giving a total of 626 different specimens (or, equivalently, images). The issues correspond to 44 different issuing authorities and span the time period between 27 BC and 244 AD, i.e., the entire duration of the Empire until the economic crisis of the third century AD, when the consistency and quality of coinage declined severely. The full list of issues included, organized by the issuing authority in alphabetical order, is shown in Table 1.

For the sake of illustration, we also include the different examples of different issues in our data set in Figures 7–10.

**Table 1.** Summary of the content of our corpus.

Issuing Authority	Number of Issues	Issues (RIC)	Time Period
Aelius	2	434, 436	136–138 AD
Antoninus Pius	5	61, 62, 111, 137, 436.	145–161 AD
Augustus	5	102, 254, 257, 270, 543	27 BC–14 AD
Caligula	2	2, 14	37–41 AD
Caracalla	3	2, 150, 658	198–217 AD
Claudius	2	10, 41	41–54 AD
Clodius Albinus	3	1, 4, 7	195–196 AD
Commodus	3	249, 251, 259	175–192 AD
Crispina	1	283	178–187 AD
Diadumenian	1	102	217–218 AD
Domitian	1	720	69–96 AD
Elagabalus	2	88, 156	218–222 AD
Faustina I (the Elder)	1	355	138–161 AD
Faustina II (Junior)	2	384, 517	147–176 AD
Galba	3	145, 167, 189	68–69 AD
Geta	2	51, 107	209–212 AD
Gordian III	3	108, 127, 130	238–244 AD
Hadrian	5	160, 169, 240, 297, 367	117–138 AD
Julia Domna	3	536, 539, 580	193–217 AD
Julia Maesa	2	268, 271	218–222 AD
Julia Mamaea	1	351	221–235 AD
Julia Paula	1	211	219–220 AD
Julia Soaemias	1	243	218–222 AD
Lucilla	1	770	164–169 AD
Lucius Verus	2	516, 576	161–169 AD
Macrinus	1	86	217–218 AD
Marcus Aurelius	4	50, 171, 426, 479	161–180 AD
Maximinus	1	23	235–238 AD
Nero	2	47, 67	54–68 AD
Nerva	2	15, 17	96–98 AD
Orbiana	1	319	225–227 AD
Otho	2	10, 12	69 AD
Pertinax	2	8, 11	193 AD
Pescennius Niger	1	64	193–194 AD
Plautilla	2	363, 369	202–205 AD
Sabina	2	395, 398	128–137 AD
Septimius Severus	6	1, 69, 106, 167, 418, 425	193–211 AD
Severus Alexander	1	146	222–235 AD
Tiberius	2	30, 224	14–37 AD
Titus	2	25, 220	79–81 AD
Trajan	4	9, 12, 243, 337	98–117 AD
Vespasian	5	2, 15, 30, 63, 99	69–79 AD
Vitellius	2	56, 109	69 AD
<b>Total</b>	<b>100</b>		<b>27 BC–244 AD</b>



**Figure 7.** Our data set entry for Aelius Denarius RIC 434—there are images of two specimens for each of the three conditions included for all denarii issues in the data set (F, VF, and XF).



**Figure 8.** Our data set entry for Antoninus Pius Denarius RIC 111—there are images of two specimens for each of the three conditions included for all denarii issues in the data set (F, VF, and XF).



**Figure 9.** Our data set entry for Octavian Augustus Denarius RIC 257—there are images of two specimens for each of the three conditions included for all denarii issues in the data set (F, VF, and XF).



**Figure 10.** Our data set entry for Commodus Denarius RIC 251—there are images of two specimens for each of the three conditions included for all denarii issues in the data set (F, VF, and XF).

### 3.4. *Caveat Scolasticus*

Before concluding, considering the general spirit of the present article, focused around the issues of research direction in the field of automated ancient coin analysis, methodological rigor, and repeatability, we would like to make note of and highlight one important issue that the community should take notice of in future work. Most directly, our observation concerns SIFT features [50], which have been used extensively in past research [13,14,26,51]. Although we note that these methods have failed to demonstrate much success, it is possible that they will find their use within differently constituted frameworks in the future. Moreover, the general point has much wider applicability.

In particular, we found that different implementations of what nominally appears to be the same deterministic algorithm of a well-known and widely used technique actually differ greatly in the output they produce. For example, we found that on the data corpus introduced in this section, the OpenCV implementation of SIFT results in approximately 1250 detections of local features per image, in contrast to the VLFeat one, which produces approximately 415. This is a huge difference (threefold, or half an order of magnitude) which undoubtedly affects any subsequent processing, representation, and learning. Thus, first and foremost when reporting their findings but also when designing new algorithms, researchers should make sure that they explicitly note the exact implementation of any technique used, no matter how standard, as well as the values of all relevant parameters.

## 4. Summary

In this paper, we addressed a number of important issues in the increasingly active research domain of application of computer vision and machine-learning-based analysis of ancient coins, which has received insufficient attention to date.

Our first contribution comes in the form of the first overview and discussion of different challenges in the field. Its aim is to clarify the intellectual landscape of automated ancient coin analysis and help direct future research efforts. Indeed, many of the discussed challenges have been scarcely investigated to date, while others have hitherto been unrecognized and remain entirely unaddressed.

Our second contribution concerns the increasingly obvious lack of standardized and appropriately curated data sets, crucial for facilitating methodological evaluation of algorithms and, in particular, the effects that coin preservation has on the performance of different methods. Indeed, until now, only one published work at all recognized the need for this kind of analysis, which, to any numismatist, would be a trivially obvious fact. Hence, we introduce the first data set to be carefully curated and collected for this purpose. We also discuss a wide range of considerations which had to be taken into account in collecting this corpus, explain our decisions, and describe its content in detail. In summary, the data set comprises 100 different coin issues, all with multiple examples in Fine, Very Fine, and Extremely Fine conditions, giving a total of over 650 different specimens (and thus images). These correspond to and depict 44 issuing authorities and span the time period of approximately 300 years, namely from 27 BC until 244 AD.

Lastly, the present article includes the first recognition of the variability in the functioning of different implementations of seemingly identical and standard 'off-the-shelf' techniques. The lack of appreciation of this fact and the lack of reporting of the details of the exact implementations used in experimental evaluations raise serious concerns as regards our understanding of the performance of different algorithms, limit insights that can be derived from the published findings, and bring forth methodological issues faced by the community. Hence, we argue for the need for greater awareness of such considerations and advise researchers to report the relevant details in the future without omission.

Our hope is that the present article should be an invaluable resource to researchers in the field, and we encourage the community to adopt the collected corpus, freely available for research purposes, as a standard evaluation benchmark.

**Author Contributions:** All authors have contributed to all aspects of the described work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ranjan, R.; Patel, V.M.; Chellappa, R. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *41*, 121–135. [[CrossRef](#)] [[PubMed](#)]
2. Viola, P.; Jones, M. Robust real-time face detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154. [[CrossRef](#)]
3. Zhou, B.; Lapedriza, A.; Xiao, J.; Torralla, A.; Oliva, A. Learning deep features for scene recognition using places database. *Adv. Neural Inf. Process. Syst.* **2014**, *1*, 487–495.
4. Niemeyer, M.; Arandjelović, O. Automatic semantic labelling of images by their content using non-parametric Bayesian machine learning and image search using synthetically generated image collages. In Proceedings of the IEEE International Conference on Data Science and Advanced Analytics, Turin, Italy, 1–3 October 2018; pp. 160–168.
5. Brzezicki, M.A.; Bridger, N.E.; Kobetić, M.D.; Ostrowski, M.; Grabowski, W.; Gill, S.S.; Neumann, S. Artificial intelligence outperforms human students in conducting neurosurgical audits. *Clin. Neurol. Neurosurg.* **2020**, *192*, 105732. [[CrossRef](#)] [[PubMed](#)]
6. Dimitriou, N.; Arandjelović, O.; Harrison, D.; Caie, P.D. A principled machine learning framework improves accuracy of stage II colorectal cancer prognosis. *NPJ Digit. Med.* **2018**, *1*, 1–9. [[CrossRef](#)] [[PubMed](#)]
7. Yue, X.; Dimitriou, N.; Arandjelović, O. Colorectal cancer outcome prediction from H&E whole slide images using machine learning and automatically inferred phenotype profiles. In Proceedings of the International Conference on Bioinformatics and Computational Biology, Honolulu, HI, USA, 18–20 March 2019.
8. Goeting, M. Seeing the world through machinic eyes: Reflections on Computer vision in the arts. In Proceedings of the Workshops of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 653–670.
9. Lang, S.; Ommer, B. Attesting similarity: Supporting the organization and study of art image collections with computer vision. *Digit. Scholarsh. Humanit.* **2018**, *33*, 845–856. [[CrossRef](#)]
10. Brown, I.; Arandjelović, O. Making Japanese ukiyo-e art 3D in real-time. *Sci* **2020**, *2*, 6. [[CrossRef](#)]
11. Jones, J.R.M. *A Dictionary of Ancient Roman Coins*; Seaby: London, UK, 1990.
12. Sutherland, C.H.V.; Carson, R.A.G. *The Roman Imperial Coinage*; Spink: London, UK, 1923; Volume 1–10.
13. Arandjelović, O. Reading ancient coins: Automatically identifying denarii using obverse legend seeded retrieval. *Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012*; Springer: Cham, Switzerland, 2012; Volume 4, pp. 317–330.
14. Arandjelović, O. Automatic attribution of ancient Roman imperial coins. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 18–13 June 2010; pp. 1728–1734.
15. Zaharieva, M.; Kappel, M.; Zambanini, S. Image based recognition of ancient coins. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, Vienna, Austria, 27–29 August 2007; pp. 547–554.
16. Aslan, S.; Vascon, S.; Pelillo, M. Two sides of the same coin: Improved ancient coin classification using Graph Transduction Games. *Pattern Recognit. Lett.* **2020**, *131*, 158–165. [[CrossRef](#)]
17. Schlag, I.; Arandjelović, O. Ancient Roman coin recognition in the wild using deep learning based recognition of artistically depicted face profiles. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2898–2906.
18. Cooper, J.; Arandjelović, O. Visually understanding rather than merely matching ancient coin images. In Proceedings of the INNS Conference on Big Data and Deep Learning, SESTRI LEVANTE, Genoa, Italy, 16–18 April 2019; pp. 330–340.
19. Conn, B.; Arandjelović, O. Towards computer vision based ancient coin recognition in the wild—Automatic reliable image preprocessing and normalization. In Proceedings of the IEEE International Joint Conference on Neural Networks, Anchorage, AK, USA, 14–19 May 2017; pp. 1457–1464.

20. Cooper, J.; Arandjelović, O. Learning to describe: A new approach to computer vision for ancient coin analysis. *Sci* **2020**, *2*, 8. [[CrossRef](#)]
21. Guo, Y.; Liu, Y.; Georgiou, T.; Lew, M.S. A review of semantic segmentation using deep neural networks. *Int. J. Multimed. Inf. Retr.* **2018**, *7*, 87–93. [[CrossRef](#)]
22. Mattingly, H. *The Roman Imperial Coinage*; Spink: London, UK, 1966; Volume 7.
23. Hartmann, C.; Hammerl, F.; Volk, W. Experimental analysis of Roman coin minting. *J. Archaeol. Sci. Rep.* **2019**, *25*, 498–506. [[CrossRef](#)]
24. Prokopov, I.; Manov, R. *Counterfeit Studios and Their Coins*; SP-P & Provias: Sofia, Bulgaria, 2005.
25. Lowe, D.G. Local feature view clustering for 3D object recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001; pp. 682–688.
26. Kampel, M.; Zaharieva, M. Recognizing ancient coins based on local features. In *Proceedings of the International Symposium on Visual Computing, Las Vegas, NV, USA, 1–3 December 2008*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 1, pp. 11–22.
27. Azad, P.; Asfour, T.; Dillmann, R. Combining Harris interest points and the SIFT descriptor for fast scale-invariant object recognition. In Proceedings of the International Conference on Intelligent Robots and Systems, St. Louis, MO, USA, 10–15 October 2009; pp. 4275–4280.
28. Wang, B.; Liang, W.; Wang, Y.; Liang, Y. Head pose estimation with combined 2D SIFT and 3D HOG features. In Proceedings of the International Conference on Image and Graphics, Qingdao, China, 26–28 July 2013; pp. 650–655.
29. Rieutort-Louis, W.; Arandjelović, O. Description transition tables for object retrieval using unconstrained cluttered video acquired using a consumer level handheld mobile device. In Proceedings of the IEEE International Joint Conference on Neural Networks, Vancouver, BC, Canada, 24–29 July 2016; pp. 3030–3037.
30. Martin, R.; Arandjelović, O. Multiple-object tracking in cluttered and crowded public spaces. In *Proceedings of the International Symposium on Visual Computing, 29 November–1 December 2010*; Springer: Berlin/Heidelberg, Germany, 2010; Volume 3, pp. 89–98.
31. Arandjelović, O. Object matching using boundary descriptors. In Proceedings of the British Machine Vision Conference, Surrey, UK, 3–7 September 2012, doi:10.5244/C.26.85. [[CrossRef](#)]
32. Arandjelović, O. Matching Objects across the Textured—Smooth Continuum. In Proceedings of the Australasian Conference on Robotics and Automation, Wellington, New Zealand, 3–5 December 2012; pp. 354–361.
33. Llamas, J.; Lerones, P.M.; Zalama, E.; Gómez-García-Bermejo, J. Applying deep learning techniques to cultural heritage images within the inception project. In Proceedings of the Euro-Mediterranean Conference; Association for Computing Machinery: New York, NY, USA, 2016; pp. 25–32.
34. Obeso, A.M.; Vázquez, M.S.G.; Acosta, A.A.R.; Benois-Pineau, J. Connoisseur: Classification of styles of Mexican architectural heritage with deep learning and visual attention prediction. In Proceedings of the International Workshop on Content-based Multimedia Indexing, Florence, Italy, 19–21 June 2017; pp. 1–7.
35. Maltezos, E.; Protopapadakis, E.; Doulamis, N.; Doulamis, A.; Ioannidis, C. Understanding historical cityscapes from aerial imagery through machine learning. In Proceedings of the Euro-Mediterranean Conference, Nicosia, Cyprus, 29 October– 3 November 2018; pp. 200–211.
36. Anwar, H.; Anwar, S.; Zambanini, S.; Porikli, F. CoinNet: Deep ancient Roman republican coin classification via feature fusion and attention. *arXiv* **2019**, arXiv:1908.09428.
37. Kampel, M.; Zambanini, S.; Schlapke, M.; Breuckmann, B. Highly detailed 3D scanning of ancient coins. In Proceedings of the International Committee of Architectural Photogrammetry Symposium, Kyoto, Japan, 11–15 October 2009; pp. 11–15.
38. Spagnolo, G.S.; Majo, R.; Carli, M.; Ambrosini, D.; Paoletti, D. Virtual gallery of ancient coins through conoscopic holography. In *Proceedings of the Optical Metrology for Arts and Multimedia*; SPIE: Bellingham, WA, USA, 2003; Volume 146, pp. 202–209.
39. Zambanini, S.; Schlapke, M.; Hödlmoser, M.; Kampel, M. 3D acquisition of historical coins and its application area in numismatics. In Proceedings of the Computer Vision and Image Analysis of Art, San Jose, CA, USA, 16 February 2010; Volume 7531, pp. 753108.
40. Aicardi, I.; Chiabrando, F.; Lingua, A.M.; Noardo, F. Recent trends in cultural heritage 3D survey: The photogrammetric computer vision approach. *J. Cult. Herit.* **2018**, *32*, 257–266. [[CrossRef](#)]
41. Makantasis, K.; Doulamis, A.; Doulamis, N.; Ioannides, M. In the wild image retrieval and clustering for 3D cultural heritage landmarks reconstruction. *Multimed. Tools Appl.* **2016**, *75*, 3593–3629. [[CrossRef](#)]

42. Voulodimos, A.; Doulamis, N.; Fritsch, D.; Makantasis, K.; Doulamis, A.; Klein, M. Four-dimensional reconstruction of cultural heritage sites based on photogrammetry and clustering. *J. Electron. Imaging* **2016**, *26*, 011013. [[CrossRef](#)]
43. Doulamis, N.; Doulamis, A.; Ioannidis, C.; Klein, M.; Ioannides, M. *Modelling of Static and Moving Objects: Digitizing Tangible and Intangible Cultural Heritage*; Springer: Cham, Switzerland, 2017; pp. 567–589.
44. Doulamis, A.; Doulamis, N.; Protopapadakis, E.; Voulodimos, A.; Ioannides, M. 4D modelling in cultural heritage. In Proceedings of the International Workshop on Advances in Digital Cultural Heritage, Madeira, Portugal, 20 February 2018; pp. 174–196.
45. France, F.G. Advanced spectral imaging for noninvasive microanalysis of cultural heritage materials: Review of application to documents in the US Library of Congress. *Appl. Spectrosc.* **2011**, *65*, 565–574. [[CrossRef](#)] [[PubMed](#)]
46. Khan, M.J.; Khan, H.S.; Yousaf, A.; Khurshid, K.; Abbas, A. Modern trends in hyperspectral image analysis: A review. *IEEE Access* **2018**, *6*, 14118–14129. [[CrossRef](#)]
47. Tonazzini, A.; Salerno, E.; Abdel-Salam, Z.A.; Harith, M.A.; Marras, L.; Botto, A.; Campanella, B.; Legnaioli, S.; Pagnotta, S.; Poggialini, F. Analytical and mathematical methods for revealing hidden details in ancient manuscripts and paintings: A review. *J. Adv. Res.* **2019**, *17*, 31–42. [[CrossRef](#)] [[PubMed](#)]
48. Fare, C.; Arandjelović, O. Ancient Roman coin retrieval: A new dataset and a systematic examination of the effects of coin grade. In Proceedings of the European Conference on Information Retrieval, Aberdeen, Scotland, UK, 8–13 April 2017; pp. 410–423.
49. Zambanini, S.; Kavelar, A.; Kampel, M. Classifying ancient coins by local feature matching and pairwise geometric consistency evaluation. In Proceedings of the International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 3032–3037.
50. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2003**, *60*, 91–110. [[CrossRef](#)]
51. Anwar, H.; Zambanini, S.; Kampel, M. Supporting Ancient Coin Classification by Image-Based Reverse Side Symbol Recognition. In Proceedings of the International Conference on Computer Analysis of Images and Patterns; Springer: Berlin, Germany, 2013; pp. 17–25.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).